Measuring Students' Critical Thinking in Online Environment: Methodology, Conceptual Framework and Tasks Typology

K.V. Tarasova, E.A. Orel

Ksenia V. Tarasova — Candidate of Sciences in Education, Deputy Head, Laboratory for Measuring New Constructs and Test Design, Centre for Psychometrics and Measurement in Education, National Research University Higher School of Economics. Address: 20 Myasnitskaya Str., 101000 Moscow, Russian Federation. E-mail: ktarasova@hse.ru (corresponding author)
Ekaterina A. Orel — Candidate of Sciences in Psychology, PhD, Head of Laboratory for Modelling and Assessing of Competencies in Higher Education, National Research University Higher School of Economics. E-mail: eorel@hse.ru

Abstract

Today critical thinking is one of the key competencies of the modern world. The abundance and availability of various information (in particular due to the spread of electronic devices and the Internet) suggest that people, regardless of age, need to be able to effectively navigate and evaluate information, draw their own conclusions and use arguments when making decisions. Research of critical thinking in terms of the possibilities of its evaluation and development began more than 60 years ago and has a wide and heterogeneous field of theoretical approaches.

In this article, we explore the theoretical possibility of measuring the complex latent construct "critical thinking in an online environment" using the Evidence-Centered Design (ECD) methodology. It allows one to build a chain of arguments that substantiate the conclusion about the level of critical thinking development in the respondent and thereby ensure the fairness of the assessment process. The measurement goes from theoretical assumptions about the nature of the construct to the search for empirical evidence — observable actions in the testing process, allowing to draw reasonable conclusions about the respondents. The result of the work is a theoretical framework for assessing critical thinking in an online environment for university students and its operationalization through the relevant observed behavior of the target audience, which allows obtaining valid data on the severity of critical thinking. This is the first step to create an instrument in Russian with a confirmed psychometric quality. The key feature of the tool is that the student does not work in a simulated environment where information sources are selected by developers, but in an open online environment, therefore, he can use any available materials to solve the task.

Keywords universal competencies; critical thinking; higher education; Evidence-Centered Design; computer-based testing assessment; online environment.

Critical thinking is one of the competencies that belong to the key skills of the XXI century [Griffin, Care, 2014]. Its significance is particularly evident when it comes to problem-solving in online environments. In order to work productively amidst the colossal amounts of information available on the Internet, the overlapping flows of data from diverse sources, and the ease of searching combined with the absence of systematic work with primary sources, it is necessary to possess skills in information analysis, making valid judgments, and establishing cause-and-effect relationships. At the same time, a person should be able to judge the reliability of information, as distorted or false information, including fake news, can cause irritation or even mislead a small community or the population of an entire country [Beer de, Matthee, 2021; Khan, 2020; Probierz, Stefanski, Kozak, 2021]. The education system faces the task of forming and evaluating critical thinking.

This study is devoted to the assessment of critical thinking among university students. The team led by I. Uglanova has been conducting similar research in the field of school education in recent years [Uglanova, Brun, Vasin, 2018; Uglanova, Orel, Brun, 2020; Uglanova, Pogozhina, 2021].

At the level of higher education, critical thinking is one of the universal competencies listed in the Federal State Educational Standard for Higher Education (FGOS VO)[1]. Universal competencies in a bachelor's degree program is a list of ten competencies that should be developed in all students at this stage of education, regardless of their specialization. Systemic and critical thinking (UC-1) is defined in this list as the ability to "search, critically analyze, and synthesize information, apply a systematic approach to problem-solving"[2]. Thus, the education system undertakes to develop critical thinking, and therefore, requires independent assessment tools to confirm the declared result.

Today, the development of critical thinking has become one of the main requests of employers to employees, including university graduates [Stepashkina, Sukhodolin, Guzhelya, 2022]. A study of job requirements collected in the O*Net system[3] showed that critical

---

[1] Ministry of Education and Science of the Russian Federation (2018) FGOS VO — Bachelor's degree: https://fgosvo.ru/fgosvo/index/24/28

[2] Critical and systemic thinking are not synonymous, however, for some reason, the compilers of the Federal State Educational Standard have combined them into a single competence.

[3] An open database that stores job descriptions and requirements for candidates: https://www.onetonline.org/

thinking, along with communication skills, is highly valued by employers in a variety of fields [Carnevale, Smith, 2013]. Critical thinking (or its analogues) is included in all universal competency models developed by leading consulting companies. Based on these models, organizations make personnel decisions — on hiring, career advancement, and providing feedback to employees. The models use different names for competencies that are similar in content, which can be attributed to critical thinking. Thus, in R. Boyatzis' model [2008], this competence appears under the names "logical thinking" and "conceptualization". Logical thinking is defined as a mental process in which a person places an event in a causal sequence based on the perception of a series of cause-and-effect relationships. Conceptualization is the process of thinking in which a person identifies or recognizes patterns in information. In L. Spencer and S. Spencer's model, the cluster of key thinking competencies includes analytical and conceptual thinking. Analytical thinking involves a systematic organization of parts of a problem or situation, conducting systematic comparisons of properties or characteristics, rational prioritization, determination of temporal sequence, and causal relationships. Conceptual thinking involves understanding a situation or problem by combining parts, and looking at the picture of events as a whole [Spencer, Spencer, 2005]. Similar competencies in content are present in other contemporary models such as "Twenty facets" (systematic thinking) [Simonenko, 2012]), WAVE ("explores options for solutions") [Kurz, 2009], Great-8 (analytical and interpretative abilities) [Bartram, 2005], and others. A substantive analysis of the mentioned competencies shows that they are all quite similar in content, focused on working with information, analysis, search for cause-and-effect relationships, and assume the ability to analyze, select, compare, interpret, and make judgments[4]. Thus, the content of the competencies included in the universal models that are used for personnel decision-making in business organizations indicates that critical thinking and related competencies are one of the most important criteria for making personnel decisions in the business environment.

Numerous diagnostic tools have been developed to assess critical thinking among university students and adult populations. However, the majority of these tools are English-language based and require high costs for adaptation and testing. There is a dearth of Russian-language assessment tools for critical thinking, and information regarding their quality and psychometric properties is largely absent [Koreshnikova, Frumin, Pashchenko, 2020].

---

[4] Gorbunova A.V. (2012) Issledovanie klyuchevyh kompetencij menedzherov vysshego i srednego zvena v Rossii [Research of key competencies of senior and middle managers in Russia] (Unpublished Master's thesis). Institute of Education of the Higher School of Economics, Moscow, Russia.

The purpose of this paper is to present a theoretical framework for assessing critical thinking in an online environment among university students as the first step towards creating a Russian-language tool with confirmed psychometric quality.

1. Approaches to the Definition of Critical Thinking
1.1. Critical Thinking in the Philosophical Tradition of Cognition Study

The research on critical thinking is based on a long philosophical tradition of cognition study. In this paradigm, the main qualities of a critically thinking person in this paradigm are reasoned judgments, purposeful thinking processes, reflection by the subject, and adherence to the rules of formal logic (for example, [Paul, Elder, 2011]). Most researchers working within the philosophical tradition agree that the presence of critical thinking can be judged by a person's ability to make a rational, conscious decision about what to do or what to believe [Ennis, 1993; Hitchcock, 2020; Norris, 1985].

R. Ennis defines critical thinking as grounded reflexive thinking aimed at determining what to believe and what not to believe [Ennis, 1993]. He identifies 10 mental operations as its components, including the assessment of the reliability of sources, the identification of various types of statements, and the skill to ask clarifying questions. Similarly, P. Facione, together with a pool of experts, developed the definition of critical thinking for use in regulatory documents and assessment of students' progress in university education. He defines critical thinking as "interpretation, analysis, evaluation, and inference, as well as explanations of the evidential, conceptual, methodological, criteriological or contextual considerations that judgment is based upon" [Facione, 1990]. According to P. Facione, the components of critical thinking include categorization of the types of utterance, evaluation, inference, explanation, and so on, totaling 6 mental operations.

In recent years, within the framework of the philosophical approach, there has been a growing popularity of generalizing concepts aimed at identifying the components of critical thinking that are common to different authors and attempting to create a consensus framework. One example of such an approach is the list of components of critical thinking proposed by E. Lai [Lai, 2011]. She identifies four components that are common to different descriptions of this construct: 1) analysis of arguments, statements, and evidence; 2) the ability to draw a conclusion using induction or deduction; 3) evaluation and judgment; 4) decision-making or problem-solving.

The model proposed by L. Liu and her colleagues [Liu, Frankel, Roohr, 2014] belongs to the same generalizing type of definitions of critical thinking. The essential difference between this model and other generalizing concepts is that it is based not only on theoretical concepts, but also on empirical results. This model identi-

fies three components of critical thinking: an analytical component (assessment of source reliability, argument relevance, search for alternative opinions and viewpoints), a synthetic component (logical inference, assessment of consequences, construction of one's own argument structure), and a general component (construction of cause-and-effect relationships and assessment of alternative explanations). This model formed the basis of *the HEIghten Critical Thinking Assessment* test, which is used to assess the critical thinking of university students.

1.2. Critical Thinking in Education Research and Practice

About 100 years ago, the first attempts to comprehend the place and role of critical thinking in education research and practice were made. The American philosopher and educator D. Dewey believed that reflexive thinking (now understood as critical thinking) should become the basis of learning [Kennedy, Fisher, Ennis, 1991]. D. Dewey defined reflexive thinking as "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends" [Dewey, 1909].

Dewey's ideas formed the basis for the development of taxonomies for educational objectives that include critical thinking and its various components. In particular, B. Bloom's taxonomy identifies six levels of educational objectives that vary in depth of mastering the material: knowledge, comprehension, application, analysis, synthesis, and evaluation [Seaman, 2011]. The last three levels are components of critical thinking. The analysis allows not only to apply specific knowledge, but also to identify patterns or algorithms suitable for problem-solving. Synthesis is a logical operation underlying the conclusion, it helps to construct a system that generalizes and explains various facts. Evaluation allows judgments to be made about the significance and value of ideas, methods, or phenomena.

In Russian pedagogy, the problems of the development of thinking at school were discussed, in particular, by P.P. Blonsky [Blonsky, 1935]. He considered that it was necessary to develop a student's critical attitude by the end of the "central part" of the school curriculum. A child acquires critical thinking when he has the opportunity to observe a teacher's reactions to rumors spread or unverified statements made by a student. A teacher can contribute to the development of a child's critical thinking by teaching him to observe, not to rush to conclusions, to be critical even of his own thoughts, to collect as many facts as possible confirming or refuting the information received, as well as by explaining the basics of causal analysis. According to Blonsky, the full development of critical thinking can be achieved by the end of school. This conclusion is consistent with the concept of J. Piaget's theory of cognitive de-

velopment [2008], according to which the transition to the stage of formal operations, representing the most complete development of thinking, begins no earlier than 12 years of age for the most advanced children [Jackson, 1965; Lovell, 1961].

D.B. Elkonin [1971] noted that by the end of primary school education, a verbal-logical type of thinking is formed, as well as a theoretical one — if structural units of theoretical generalization according to V.V. Davydov [1996] were used in teaching. These types of thinking precede the formation of critical thinking and serve as its foundation.

1.3. Critical Thinking in Psychology

The psychological tradition of the study of critical thinking, as opposed to the philosophical and educational traditions, focuses on the study of the cognitive processes and the behavior of a critically thinking subject in real-life situations, rather than in ideal conditions [Lai, 2011; Sternberg, 1986]. The definitions of critical thinking proposed by cognitive psychologists are based on observable behavior as evidence of critical thinking. Environmental factors and personal characteristics of the thinking subject are also taken into account. The observable evidence and manifestations of critical thinking may not fully cover the entire construct of critical thinking, and may only represent a part of it.

One of the most well-known psychological operationalization of critical thinking was proposed by R. Sternberg. He provides the following definition: "critical thinking comprises the mental processes, strategies and representations people use to solve problems, make decisions and learn new concepts" [Sternberg, 1986]. R. Sternberg identifies the following components of critical thinking:

- the metacomponent — high-order processes related to planning what needs to be done, monitoring progress, and evaluating what has been done;
- the component of cognitive operation — processes that "serve" the instructions received from the metacomponent, such as induction, deduction, spatial visualization, and others;
- the component of knowledge — processes used for understanding new concepts and procedures.

A significant aspect of defining critical thinking is the statement made by E.V. Ilyenkov, who noted that "thinking <...> is nothing but the ability (skill) to deal intelligently with each subject, i.e. in accordance with its own nature" [Ilyenkov, 2002. P. 86]. Based on the works of E.V. Ilyenkov [1974; 1979], V.V. Davydov developed the content of the concept of "dialectical thinking", which is close to the construct discussed in the paper. According to Davydov, dialectical thinking is a type of thinking that analyzes the development of the

whole on the basis of internal contradiction. To initiate a thought process in each specific case, it is necessary to identify the initial key relation (contradiction) that generates the diversity of content as it unfolds [Davydov, 1972].

In Russian psychology, there are several concepts similar to "critical thinking", but not identical to it. One of these concepts is the criticality of thinking.

Criticality is one of the properties of normal mental activity, the ability to realize one's mistakes, the ability to evaluate one's thoughts, weigh the arguments "for" and "against", and subject one's hypotheses to thorough testing [Rubinstein, 2002; Teplov, 1946]. According to B.V. Zeigarnik [1986], criticality consists of the ability to act thoughtfully, to verify and correct one's actions in accordance with the conditions of reality. Impairments in critical thinking are a hallmark of a mental disorder that manifest in the loss of control over intellectual processes, therefore they have been mainly studied experimentally in clinical psychology.

From the given definitions of critical thinking, it is evident that this concept as understood by B.M. Teplov, L.S. Rubinstein, and B.V. Zeigarnik is similar to our understanding of critical thinking. Criticality is characterized by a focus on the subject of one's own mental activity. L.S. Rubinstein wrote that "the ability to realize one's mistake is a privilege of thought as a conscious process"; "criticality is an essential feature of a mature mind. An uncritical, naive mind easily accepts any coincidence as an explanation, and the first solution that comes to mind as final. The critical mind carefully weighs all the arguments "for" and "against" its hypotheses and subjects them to thorough testing" [Rubinstein, 2002]. Thus, in psychological as well as in philosophical theories of thinking, both cognitive and regulatory aspects of the thought process are taken into account.

Most standardized methods for assessing critical thinking are based on a philosophical approach, including *California Critical Thinking Skills Test* (CCTST), *Cornell Critical Thinking Test* (CCTT), *Ennis-Weir Critical Thinking Essay Test* (EWCTET), *HEIghten Critical Thinking Assessment*. Philosophers engaged in the study of critical thinking examine it from the perspective of ideal properties, which allows for a clear theoretical structure of the studied construct. For this reason, the developers of assessment tools predominantly rely on this tradition of critical thinking research. In the educational tradition, the greatest attention is paid to developing critical thinking, but the theoretical framework of the construct remains less developed. Definitions of critical thinking within the framework of the psychological approach are distinguished by deep theoretical elaboration and orientation to the description and explanation of the deep processes that determine critical thinking, but are less oriented towards creating mass assessment tools.

2. Measurement Tools for Critical Thinking

With the growing demand for critical thinking skills in modern society, the number of measurement tools is also increasing. In this regard, let us consider the tools created for university and college students, with the confirmed validity and reliability of the obtained results.

One of the first standardized tests of critical thinking, widely used until now, is the *Watson-Glaser Critical Thinking Appraisal tool* (WGCTA) [Watson, Glaser, 1980]. It is based on the concept of critical thinking as the ability to identify and analyze problems, as well as to search for and evaluate the necessary information in order to come to the desired conclusion. The first version of the test appeared in 1960, and later it was repeatedly changed and modified. In 2011, a computer adaptive version of the test was developed. The researchers note the high discriminative power of the tasks, a large task bank, a high level of reliability of complete test forms, and a significant level of predictive validity. Nevertheless, the test is not without its drawbacks, including low design validity associated with deficiencies in task instructions [Possin, 2014], and insufficient internal consistency of some subscales [Bernard et al., 2008].

R. Ennis, a well-known researcher of critical thinking within the philosophical approach, played a crucial role at the inception of the *Cornell Critical Thinking Test* (CCTT) [Ennis, Millman, Tomko, 2005] and the *Ennis-Weir Critical Thinking Essay Test* (EWCTET) [Ennis, Weir, 1985]. The reliability coefficients of the Cornell test range from 0.67 to 0.90. The Ennis-Weir test essay consists of 9 open-ended questions and focuses on assessing general argumentation skills. Unlike multiple-choice tests, it allows students to justify their answers. The CCTT is based on the conception of critical thinking as a reflexive and rational inference focused on what to believe or what to do [Ennis, 1993], while the authors of the EWCTET emphasize the creative aspect of critical thinking, taking into account Ennis's developments and defining the construct as a person's ability to evaluate an argument and formulate a written response. The main disadvantage of the essay test is the need to involve experts to evaluate open-ended responses.

The *California Critical Thinking Skills Test* (CCTST) is based on the works of another representative of the philosophical tradition in the study of critical thinking — P. Facione. According to the conceptualization formulated by a group of 46 national experts, critical thinking is a purposeful, self-regulating judgment that leads to interpretation, analysis, evaluation, and inference, as well as explanation of evidential, conceptual, methodological, criteriological, or contextual considerations upon which the judgment is based [Facione, 1990]. Studies have confirmed the high reliability of the CCTST, which ranges from 0.7 to 0.84 depending on the version of the test [Behar-Horenstein, Niu, 2011], however, the disadvantage of

the test is the ambiguity of the formulations: depending on their interpretation, there may be several correct answers, which negatively affect the quality of measurement results [Fawkes et al., 2005].

The *Council for Aid to Education* (CAE) in the USA has developed the *Collegiate Learning Assessment* (CLA) tool, as well as its improved version CLA+ [Aloisi, Callaghan, 2018] for assessing critical thinking in an activity-based format. The tool can also be used for both individual assessment in order to provide students with feedback on their level of development in critical thinking and written communication skills, and to assess the effectiveness of the faculty/university curricula for accreditation and for reporting. At the same time, researchers note insufficient reliability of the test: at the individual level, it is unsuitable for drawing conclusions about changes in students' critical thinking levels [Ibid.], and at the institutional level, it cannot be used for high-stakes decision-making [Steedle, 2012].

The *Educational Testing Service* (ETS) has recently developed the *HEIghten Critical Thinking Assessment*, a computer-based test designed for university students. The authors consider critical thinking as a complex construct that cannot be measured holistically, but rather by evaluating its main components — analysis and synthesis of information. The test is designed to determine the level of formation of critical thinking among students, to identify strengths and weaknesses of educational programs and opportunities for their improvement [Liu, Frankel, Roohr, 2014].

The international iPAL project [Zlatkin-Troitschanskaia et al., 2018] is aimed at developing a new generation of assessment tools for higher education and professional activity, including those designed to measure critical thinking. The iPAL uses the experience gained in the development of the CLA and the developments of another international project for assessing the quality of higher education — AHELO [Tremblay, 2013]. Within the framework of the iPAL project, based on a holistic approach and ECD (*Evidence-Centered Design*) methodology [Mislevy, Almond, Lukas, 2003], tools are being developed using scenario-based tasks (all scenarios are created with the possibility of cross-cultural comparisons), modeling real-life situations.

Despite the use of innovative methodologies in the development of assessment tools, the quality of the created scenarios depends on the skill of the task developers and on the evaluation procedure itself. The creators of these methodologies have realized that an open Internet environment is necessary to demonstrate the formed critical thinking skills. The next step in the development of the iPAL project was the *Critical Online Reasoning Assessment* (CORA) tool [Nagel et al., 2020]. The respondent receives a question with an ambiguous or controversial social context, in order to answer, he needs to find information on the topic in an open online environ-

ment, formulate and justify his point of view on this issue, supporting it with links to the sources found. The main difference of this tool is that student works in an open online environment and can use any available materials to solve the task, instead of a simulated environment where information sources are selected by developers. In addition to the respondent's answers to the questions posed, the process of searching for the answer is analyzed, including behavior strategy in the online environment, site selection as sources of information, data on search queries, site addresses, and time spent on them. Such tasks are maximally close to both life-based and educational real practice of students, however, they pose serious psychometric and technical challenges to developers related to data collection, storage, and processing. Moreover, the use of such assessment tools is difficult due to the need to involve experts in task checking, which inevitably slows down the assessment process.

Table 1 presents a systematized overview of the described critical thinking assessment tools.

Table 1. **Measurement Tools for Critical Thinking**

Title of the Measurement Tool	Components of Critical Thinking Included in Operational Framework	Target Audience	Format	Tasks	Time Duration
Watson–Glaser Critical Thinking Appraisal tool (WGCTA)	Logical reasoning, recognition of assumptions, level of deduction, interpretation and evaluation of arguments	Used in various fields, including measuring critical thinking in business environments for different target audiences, such as higher education institutions (for student evaluation and career guidance), private and public organizations (for candidate selection, employee assessment, and predicting job performance)	Paper-and-pencil form/ computer-based form	80 (multiple choice)	60 minutes
California Critical Thinking Skills Test (CCTST)	Interpretation, analysis, assessment, inference, deductive reasoning and inductive reasoning	College students. Currently it is also used to evaluate undergraduate students and gifted high school students	Paper-and-pencil form/ computer-based form	34 (multiple choice)	45 minutes
Cornell Critical Thinking Test (CCTT)	Form X: Analysis, Deduction, Reliability (Trustworthiness), and Identification of Assumptions. Form Z: Induction, Deduction, Reliability (Trustworthiness), Identification of Assumptions, Semantics, Definition, and Prediction for Experiment Planning.	Form X is intended for secondary and high school students. Form Z is designed for advanced high school students, college and university students, and adult target audience	Paper-and-pencil form/ computer-based form	Form X: 71 (multiple choice). Form Z: 52 (multiple choice)	50 minutes
Ennis–Weir Critical Thinking Essay Test (EWCTET)	Formulating one's own perspective, identifying causes and assumptions, stating one's own viewpoint, providing substantial	Students	Essay	9 open-ended questions	40 minutes

Title of the Measurement Tool	Components of Critical Thinking Included in Operational Framework	Target Audience	Format	Tasks	Time Duration
	reasons, identifying other possibilities, and responding adequately to shortcomings				
HEIghten Critical Thinking Assessment	The analytical component of critical thinking involves assessing the reliability of sources, relevance of arguments, and searching for alternative opinions and perspectives. The synthetic component involves logical inference, assessment of consequences, constructing one's own argument structure, and so on. The general component involves establishing cause-and-effect relationships and evaluation of alternative explanations	Students in different forms of education, including traditional, blended, and online learning	Computer-based form	The tasks involve analyzing text fragments of varying lengths to answer questions and identify arguments in favor of or against certain position	45 minutes
Critical Online Reasoning Assessment (CORA)	Skills of critical selection and evaluation of online sources and information, as well as their use for making and justifying fact-based decisions	Students	Computer-based form	5 open-ended questions	60 minutes
Collegiate Learning Assessment+ (CLA+)	Reasoning, evaluation, and critical analysis of arguments	Students	Computer-based form	Open-ended questions + multiple choice	90 minutes
Halpern Critical Thinking Assessment (HCTA)	Reasoning, analysis of arguments, hypothesis testing, use of probability and uncertainty concepts, and decision-making skills	Students	Computer-based form	Open-ended questions + multiple choice	60 minutes

Such a diversity of tools may raise the question of why new tests should be developed when existing ones can be adapted. Most assessment tools do not provide an opportunity to qualitatively assess critical thinking primarily due to the selected task format: multiple choice tasks do not allow for testing complex skills that are components of critical thinking and are likely to reflect irrelevant constructs [Liu, Frankel, Roohr, 2014]. Furthermore, none of the listed tools are open, unlike, for example, many psychological scales that are fully published in scientific journals. In these conditions, the time and financial costs of adapting an existing tool — translation, matching the theoretical framework and tasks to the cultural context, testing (if necessary, not just once), ensuring the psychometric quality of the tool meets the standard requirements — are practically the same as the costs of developing a new tool [Amer-

ican Educational Research Association, 2018; Baturin et al., 2015]. Moreover, the possibilities for using an adapted test are often limited by copyright holders. The methods of remuneration for using the test may vary, but quite often the scheme of "payment for one testing" or "payment for one report"[5] is applied, and in this case, using adapted tools in mass monitoring studies or testing students for their portfolio becomes very expensive. Additionally, in order to fulfill the requirement of equivalence of versions, copyright holders usually severely restrict the possibilities for improving the adapted tool, so the identified shortcomings are retained in all versions of the test. Thus, the adaptation of existing tools is no less expensive than the development of new ones, but at the same time significantly limits the possibilities for using a ready-made test.

3. Critical Thinking in the Online Environment: Operationalization and Assessment Tool

The literature analysis reveals that critical thinking is a complex and multidimensional latent construct. Similar elements of critical thinking are distinguished in foreign and domestic traditions, even if different terms are used to describe them.

Creating a measurement tool and operationalizing this construct requires combining philosophical and psychological approaches, since critical thinking should be evaluated through the relevant observable behavior of the participant, and such an assessment involves more than just listing the types and ways of behavior characteristic of a critically thinking person. *Performance tasks*, in which the respondent should perform certain actions, are considered as more suitable for assessing critical thinking than traditional multiple-choice tasks, since *performance tasks* present a problem in a specific context and assume answers similar to those required in a person's professional activity and everyday life [Braun, Kirsch, Yamamoto, 2011; Messick, 1994]. *Performance tasks* types describe continuous actions that unfold over time, just as they occur in real life, rather than isolated components of those actions [Braun, Kirsch, Yamamoto, 2011; Lane, Stone, 2006; Zlatkin-Troitschanskaia, Shavelson, 2019]. Additionally, an open online environment is recommended for the study of critical thinking, where the respondent is not limited by simulation resources and task developer skill.

Based on a holistic approach and the *Evidence-Centered Design* (ECD) methodology [Mislevy, Almond, Lukas, 2003], the authors are developing a tool to measure critical thinking in an online environment using *performance tasks*. Critical thinking is considered as the ability of a university student to analyze statements, assumptions,

---

[5] An example of such pricing can be found on the website of the "Humanitarian Technologies" laboratory — one of the leading commercial companies in Russia engaged in the development of tests for business purposes: https://ht-lab.ru/news/5805/

and arguments, build causal relationships, select logical and persuasive arguments, find explanations, draw conclusions, and form their own position when solving tasks in an online environment, including in an open digital environment (with access to the Internet and subsequent collection of log data and event logging). The test content is not related to the training direction — all students receive the same set of tasks that are not dependent on their educational program. One of the tasks facing the developers is to link the level of students' critical thinking with how they work with the information sources viewed during the completion of tasks, as well as their current sociocultural and technological learning environment. Subsequently, integration into the research of additional parameters, including students' attitudes and beliefs, and the level of their general intellectual development, is possible.

In creating the conceptual framework, we relied on both philosophical and psychological approaches. Within the construct of "critical thinking in an online environment", the components were identified that made it possible to present an integrative assessment model in which *critical thinking* skills, studied within the framework of a philosophical approach [Liu, Frankel, Roohr, 2014], with *critical online reasoning* skills. Drawing on the results of psychological research on critical thinking, the model presents the observable behavior of the target group as evidence of critical thinking, as well as environmental factors and age characteristics of the target audience. The construct model is presented in Table 2.

Table 2. **Theoretical Framework of the Measurement Tool for Critical Thinking in the Online Environment**

Component of Critical Thinking	Evidence	Observable Behavior	Product of Activity
Analysis: The respondent evaluates and analyzes the evidence and arguments, as well as the context of their application. The analysis allows for the identification of the relationship between information elements and the assessment of their quality, such as determining the reliability of facts, identifying the strengths and weaknesses of arguments, and evaluating their relevance to the given task	Evaluation of evidence	Categorizes arguments into different contexts	Distributes arguments to the appropriate contexts
		Evaluates the relevance of information	Evaluates information from the source(s) in terms of its degree of relevance
		Evaluates the competence of information sources	Evaluates the sources based on their degree of competence
		Evaluates the authority of information sources	Evaluates the sources based on their degree of authority
		Identifies cognitive biases in the presented evidence	Selects all relevant biases from the provided list
		Evaluates the relevance of information for the conclusion	Evaluates the presented information in terms of its degree of relevance
		Evaluates the accuracy of information	Evaluates information from the source(s) in terms of its degree of accuracy

Component of Critical Thinking	Evidence	Observable Behavior	Product of Activity
	Analysis and evaluation of arguments	Analyzes the structure of an argument	Accurately identifies explicit premises and hidden assumptions
			Identifies linguistic cues
			Identifies premises in the text
			Identifies conclusions in the text
			Identifies intermediate steps in the argumentation
		Evaluates the structure of an argument (persuasiveness/lack of persuasiveness of the argument from the perspective of its structure and interrelationships between parts of the argument)	Evaluates the persuasiveness of the argumentation
			Evaluates the logical correctness of the argument
			Points out structural shortcomings that may be present in invalid arguments
		Identifies different categories of information in the text	Determines information that can be used as an argument
		Determines insufficiency of information in the argumentation	Draws a conclusion about the sufficiency of information in the argumentation
Synthesis: the respondent makes logically correct and true conclusions and considers their consequences. Synthesis includes formulating conclusions and understanding their consequences.	Developing a conclusion	Based on the presented information for argumentation, the respondent reaches a clear judgment ("for" or "against")	Makes inferences without committing logical fallacies
		Develops valid conclusions	Writes or collects valid conclusions from the premises that support a certain position
		Develops true conclusions	Selects true premises
			Formulates a true conclusion
		Identifies alternative conclusions	Identifies alternative valid/true conclusions
			Determines the context in which a conclusion ceases to be true
	Understanding the consequences	Determines the consequences of the conclusion made in different contexts	Determines the consequences of the conclusion in different contexts
		Identifies limitations of the conclusion	Modifies the premises such that the conclusion ceases to be valid
			Modifies the premises such that the conclusion ceases to be true
Establishing cause-and-effect relationships	Establishes cause-and-effect relationships	Evaluates cause-and-effect relationships	Forms the judgment about the accuracy of the cause-and-effect chain or relationship
		Provides explanations	Explains the presented facts, answering the questions "why?" (determining causes) and "what for?" (determining effects)

The ECD methodology [Oliveri, Mislevy, 2019] allows us to move from a general construct to variables upon which the test tasks are based. ECD provides a solid foundation for assessment, allowing us to gather as much evidence as possible that the conclusion drawn about the respondent's level of proficiency in the evaluated construct based on observations and analysis of their activity during task completion reflect reality. This approach is most relevant for measuring complex constructs, since it does not require one-dimensional measurement and allows us to model the relationships that reflect their complex nature. Following the ECD methodology, in order to create a model of critical thinking in a digital environment, evidence of the manifestation of the construct is described, the relevant observable respondent behavior in task-solving, and activity products — a result of actions during testing process that can be recorded in order to form an understanding of the level of critical thinking. Further, the task forms are proposed in which these activity products can be recorded in both closed and open digital environments (Table 3).

Table 3. **Proposed Forms of Tasks**

Task Form	Description
Selecting text fragments	Task requires the respondent to select elements of the text in accordance with the instructions
Statement Selection	From a group of statements, the respondent selects those that together or separately fulfill the given role
Short Constructed Response	The respondent must answer a question presented in text, graphic, or other form in his own words
Essay	Based on provided materials, the respondent writes an essay on a given topic in which he evaluates the arguments presented in support of specific conclusions or creates his own argument in support of a particular position
Multiple Choice with One or Several Correct Options	The respondent selects one or several answer options from a provided list. He may be required to select a certain number of answers or select all that he finds suitable. The number of proposed options may vary
Text Editing	The changes made by the respondent to the provided product are evaluated. An example is editing the text with consideration for a changed audience
Classification	Distribution of text fragments into categories
Comparison	Grouping the elements according to specific characteristics or principles

4. Conclusion   One of the important challenges in measuring critical thinking (and other complex latent constructs) is the need to constantly update the context and bring the tools closer to real life. The use of a current online context for learning and assessment poses new challenges for the researcher, both in data collection and processing: it is

necessary to simultaneously process and then store large amounts of data — not only students' responses, but also the collateral information collected during testing, including data on students' behavior in the online environment. Implementation of the ECD methodology requires the use of complex mathematical models — only with their help it is possible to demonstrate how the recorded behavior of students during testing is collected as evidence of critical thinking.

The measurement of critical thinking in higher education faces not only methodological and technical problems, but also difficulties in administering the research and utilizing the results. Firstly, student motivation poses a challenge as critical thinking, although expected as a learning outcome, is not a separate subject and therefore does not require a separate evaluation. In this case, the motivation to take the test is significantly reduced, which inevitably affects the results. The second difficulty, directly related to the first, is the integration of the evaluation procedure into the educational process: what should be its status within the curriculum? Despite some positive examples, a definitive solution that satisfies all parties involved in the educational process has not yet been found.

However, despite these challenges, the methodology presented in the paper allows for the creation of a modern tool for measuring critical thinking that is suitable for monitoring evaluation. Within the developed conceptual framework, each component of the "critical thinking in an online environment" construct has a wide range of behavioral manifestations and potential activity outcomes (i.e., the results of student actions when performing critical thinking tasks). This approach enables the creation of various task scenarios that are closely aligned with the respondent's real-life experiences, thereby increasing their motivation to complete the tasks and potentially improving the collection of reliable and valid diagnostic information.

References

Aloisi C., Callaghan A. (2018) Threats to the Validity of Peer Learning Assessment (CLA +) as an Indicator of Critical Thinking Skills and the Implications for Knowledge. *Pedagogy of Higher Education*, vol. 3, no 1, pp. 57–82. doi:10.108 0/23752696.2018.1449128

American Educational Research Association (2018) *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Bartram D. (2005) The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology*, vol. 90, no 6, pp. 1185–1203. doi:10.1037/0021-9010.90.6.1185

Baturin N., Vuchetich E., Kostromina S. et al. (2015) Rossijskij standart testirovaniya personala (vremennaya versiya, sozdannaya dlya shirokogo obsuzhdeniya v 2015 g.) [Russian Standard for Personnel Testing (Interim Version, Designed for a Discussion)]. *Organizational Psychology*, vol. 5, no 2, pp. 67–138.

Beer de D., Matthee M. (2021) Approaches to Identify Fake News: A Systematic Literature Review. *Integrated Science in Digital Age 2020* (ed. T. Antipova), Cham: Springer, pp. 13–22. doi:10.1007/978-3-030-49264-9_2

Behar-Horenstein L.S., Niu L. (2011) Teaching Critical Thinking Skills in Higher Education: A Review of the Literature. *Journal of College Teaching & Learning*, vol. 8, no 2, pp. 25–41. doi:10.19030/tlc.v8i2.3554

Bernard R.M., Zhang D., Abrami P.C., Sicoly F., Borokhovski E., Surkes M. A. (2008) Exploring the Structure of the Watson–Glaser Critical Thinking Appraisal: One Scale or Many Subscales? *Thinking Skills and Creativity*, vol. 3, no 1, pp. 15–22. doi:10.1016/j.tsc.2007.11.001

Blonskiy P.P. (1935) *Razvitie myshleniya shkol'nika* [The Development of the Schoolchild's Thinking]. M.: Uchpedgiz.

Boyatzis R. (2008) *Kompetentny menedzher. Model' effektivnoj raboty* [The Competent Manager]. Moscow: HIPPO.

Braun H., Kirsch I., Yamamoto K. (2011) An Experimental Study of the Effects of Monetary Incentives on Performance on the 12th-Grade NAEP Reading Assessment. *Teachers College Record*, vol. 113, no 11, pp. 2309–2344. doi:10.1177/016146811111301101

Carnevale A.P., Smith N. (2013) Workplace Basics: The Skills Employees Need and Employers Want. *Human Resource Development International*, vol. 16, no 5, pp. 491–501. doi:10.1080/13678868.2013.821267

Davydov V.V. (1972) *Vidy obobshchenij v obuchenii* [Types of Generalizations in Learning]. M.: Pedagogika.

Davydov V.V. (1996) *Teoriya razvivayushchego obucheniya* [Theory of Developmental Learning]. Moscow: Intor.

Dewey J. (1909) *Moral Principles in Education*. Boston: Houghton Mifflin.

El'konin D.B. (1971) K probleme periodizatsii psikhicheskogo razvitiya v detskom vozraste [On the Problem of Periodization of Mental Development in Childhood]. *Voprosy Psychologii*, no 4, pp. 6–20.

Ennis R.H. (1993) Critical Thinking Assessment. *Theory into Practice*, vol. 32, no 3, pp. 179–186. doi:10.1080/00405849309543594

Ennis R.H., Millman J., Tomko T.N. (2005) *Cornell Critical Thinking Tests*. Seaside, CA: The Critical Thinking Co.

Ennis R.H., Weir E.E. (1985) *The Ennis–Weir Critical Thinking Essay Test: An Instrument for Teaching and Testing*. Pacific Grove, CA: Midwest Publications.

Facione P. (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. ERIC Doc. No ED 315 423*. Fullerton, CA: California State University.

Fawkes D., O'Meara B., Weber D., Flage D. (2005) Examining the Exam: A Critical Look at the California Critical Thinking Skills Test. *Science & Education*, vol. 14, no 2, pp. 117–135. doi:10.1007/S11191-005-6181-4

Griffin P., Care E. (eds) (2014) *Assessment and Teaching of 21st Century Skills: Methods and Approach*. Dordrecht: Springer.

Hitchcock D. (2020) Critical Thinking. *The Stanford Encyclopedia of Philosophy* (ed. E.N. Zalta), Stanford, CA: Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/fall2020/entries/critical-thinking/ (accessed 2 September 2022).

Il'enkov E.V. (2002) *Shkola dolzhna uchit' myslit'* [School Should Teach Thinking]. Moscow: Russian Academy of Education, Moscow Psychological and Social Institute.

Il'enkov E.V. (1979) *Dialekticheskoe protivorechie* [Dialectical Contradiction]. Moscow: Politizdat.

Il'enkov E.V. (1974) *Dialekticheskaya logika* [Dialectical Logic]. Moscow: Politizdat.

Jackson S. (1965) The Growth of Logical Thinking in Normal and Subnormal Children. *British Journal of Educational Psychology*, vol. 35, no 2, pp. 255–258.

Kennedy M., Fisher M.B., Ennis R.H. (1991) Critical Thinking: Literature Review and Needed Research. *Educational Values and Cognitive Instruction: Implications for Reform* (eds L. Idol, B. Fly Jones), Hillsdale, NJ: Lawrence Erlbaum, pp. 11–40.

Khan U. (2020) Developing Critical Thinking in Student Seafarers: An Exploratory Study. *Journal of Applied Learning & Teaching*, vol. 3, Sp. Iss. 1, pp. 40–50. doi:10.37074/jalt.2020.3.s1.15

Koreshnikova Yu.N., Froumin I.D., Pashchenko T.V. (2020) Bar'ery dlya sozdaniya pedagogicheskikh uslovij razvitiya kriticheskogo myshleniya v rossijskikh vuzakh [Barriers to Creating Pedagogical Conditions for the Development of Critical Thinking in Russian Universities]. *Pedagogika*, vol. 84, no 9, pp. 45–54.

Kurz R. (2009) The Structure of Work Effectiveness as Measured through the Saville Consulting Wave® Performance 360 'BA-G'Model of Behaviour, Ability and Global Performance. *Assessment & Development Matters*, vol. 1, no 1, pp. 15–18.

Lai E.R. (2011) *Critical Thinking: A Literature Review Research Report*. London: Parsons Publishing.

Lane S., Stone C. (2006) Performance Assessment. *Educational Measurement* (ed. R.L. Brennan), Westport, CT: American Council on Education and Praeger, pp. 387–431.

Liu O.L., Frankel L., Roohr K.C. (2014) Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. *ETS Research Report Series*, no 2014 (1), pp. 1–23. doi:10.1002/ets2.12009

Lovell K. (1961) A Follow-Up Study of Inhelder and Piaget's. The Growth of Logical Thinking. *British Journal of Psychology*, vol. 52, no 2, pp. 143–153.

Messick S. (1994) The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher*, vol. 23, no 2, pp. 13–23. doi:10.3102/0013189X023002013

Mislevy R.J., Almond R.G., Lukas J.F. (2003) A Brief Introduction to Evidence-Centered Design. *ETS Research Report Series*, no 2003 (1), pp. i–29.

Nagel M.-T., Schäfer S., Zlatkin-Troitschanskaia O. et al. (2020) How Do University Students' Web Search Behavior, Website Characteristics, and the Interaction of Both Influence Students' Critical Online Reasoning. *Frontiers in Education*, no 5, Article no 565062. doi:10.3389/feduc.2020.565062

Norris S.P. (1985) Synthesis of Research on Critical Thinking. *Educational Leadership*, vol. 42, no 8, pp. 40–45.

Oliveri M.E., Mislevy R.J. (2019) Introduction to "Challenges and Opportunities in the Design of 'Next-Generation Assessments of 21st Century Skills'". *International Journal of Testing. Special Issue*, vol. 19, no 2, pp. 97–102. doi:10.1080/15305058.2019.1608551

Paul R., Elder L. (2011) *Critical Thinking: Tools for Taking Charge of Your Learning and Your Life*. Upper Saddle River, NJ: Prentice Hall.

Piaget J. (2008) Rech' i myshlenie rebenka [The Language and Thought of the Child]. Moscow: Rimis.

Possin K. (2014) Critique of the Watson-Glaser Critical Thinking Appraisal Test: The More You Know, the Lower Your Score. *Informal Logic,* vol. 34, no 4, pp. 393–416. doi:10.22329/il.v34i4.4141

Probierz B., Stefański P., Kozak J. (2021) Rapid Detection of Fake News Based on Machine Learning Methods. *Procedia Computer Science*, vol. 192, pp. 2893–2902. doi:10.1016/j.procs.2021.09.060

Rubinstein L.S. (2002) Osnovy obshchej psikhologii [Fundamentals of General Psychology]. Saint Petersburg: Piter.

Seaman M. (2011) Bloom's Taxonomy: Its Evolution, Revision, and Use in the Field of Education. *Curriculum and Teaching Dialogue*, vol. 13, no 1–2, pp. 29–131A.

Simonenko S.I. (2012). Model' effektivnogo rukovoditelya v ramkakh kontseptsii dinamicheskogo liderstva [Model of Effective Manager in Frame of Dynam-

ic Leadership Concept]. *Izvestiya of Saratov University. Philosophy. Psychology. Pedagogy*, vol. 12, no 4, pp. 90–96.

Spencer L.M., Spencer S.M. (2005) Kompetentsii na rabote [Competence at Work. Models for Superior Performance]. Moscow: HIPPO.

Steedle J.T. (2012) Selecting Value-Added Models for Postsecondary Institutional Assessment. *Assessment & Evaluation in Higher Education*, vol. 37, no 6, pp. 637–652. doi:10.1080/02602938.2011.560720

Stepashkina E., Sukhodoev A., Gudgelya D. (2022) *Issledovanie profilya nadprofessional'nykh kompetentsij, vostrebovannykh vedushchimi rabotodatelyami pri prieme na rabotu studentov i vypusknikov universitetov i molodykh spetsialistov* [The Research on the Essential Range of Soft Skills Enquired by Leading Employers during the Process of Recruitment of University Graduates and Young Professionals]. Moscow: HSE.

Sternberg R.J. (1986) *Critical Thinking: Its Nature, Measurement, and Improvement. ERIC no ED272882.* Available at: https://files.eric.ed.gov/fulltext/ED272882.pdf (accessed 2 September 2022).

Teplov B. (1946) *Psikhologiya* [Psychology]. Moscow: Uchpedgiz.

Tremblay K. (2013) OECD Assessment of Higher Education Learning Outcomes (AHELO): Rationale, Challenges and Initial Insights from the Feasibility Study. *Modeling and Measuring Competencies in Higher Education* (eds S. Blömeke, O. Zlatkin-Troitschanskaia, Ch. Kuhn, Ju. Fege), Rotterdam, Boston, Taipei: Sense, pp. 113–126.

Uglanova I., Brun I., Vasin G. (2018) Metodologiya Evidence-Centered Design dlya izmereniya kompleksnykh psikhologicheskikh konstruktov [Evidence-Centered Design Method for Measuring Complex Psychological Constructs]. *Journal of Modern Foreign Psychology*, vol. 7, no 3, pp. 18–27. doi:10.17759/jmfp.2018070302

Uglanova I.L., Orel E.A., Brun I.V. (2020) Izmerenie kreativnosti i kriticheskogo myshleniya v nachal'noj shkole [Measuring Creativity and Critical Thinking in Primary School]. *Psikhologicheskij zhurnal*, vol. 41, no 6, pp. 96–107. doi:10.31857/S020595920011124-2

Uglanova I. L., Pogozhina I. N. (2021) Chto mozhet predlozhit' novaya metodologiya otsenki myshleniya shkol'nikov sovremennomu obrazovaniyu [What the New Measure of Thinking in School Students Has to Offer to Contemporary Education]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 4, pp. 8–34. https://doi.org/10.17323/1814-9545-2021-4-8-34

Watson G., Glaser E.M. (1980) *Watson–Glaser Critical Thinking Appraisal Manual*. San Antonio, TX: Psychological Corporation.

Zeigarnik B.V. (1986) *Patopsikhologiya* [Pathopsychology]. Moscow: Moscow University Publishing House.

Zlatkin-Troitschanskaia O., Shavelson R.J. (2019) Advantages and Challenges of Performance Assessment of Student Learning in Higher Education. *British Journal of Educational Psychology*, vol. 89, no 3, pp. 413–415. doi:10.1111/bjep.12314

Zlatkin-Troitschanskaia O., Toepper M., Pant H.A., Lautenbach C., Kuhn C. (2018) *Assessment of Learning Outcomes in Higher Education: Cross-National Comparisons and Perspectives*. Cham: Springer.